



RESEARCH ARTICLE

10.1002/2017WR020767

Special Section:

Engagement, Communication,
and Decision-Making Under
Uncertainty

Key Points:

- New and existing sensitivity analysis methods are combined to create a robust approach for identifying parameter importance and interactions
- Computational demands are reduced and insights increased by using global and derivative-based methods
- Demonstration using the FUSE hydrological modeling framework highlights the new graphical presentation techniques

Supporting Information:

- Supporting Information S1

Correspondence to:

E. Borgonovo,
emanuele.borgonovo@unibocconi.it;
O. Rakovec,
oldrich.rakovec@ufz.de

Citation:

Borgonovo, E., X. Lu, E. Plischke,
O. Rakovec, and M. C. Hill (2017),
Making the most out of a hydrological
model data set: Sensitivity analyses to
open the model black-box, *Water
Resour. Res.*, 53, 7933–7950,
doi:10.1002/2017WR020767.

Received 14 MAR 2017

Accepted 9 AUG 2017

Accepted article online 21 AUG 2017

Published online 19 SEP 2017

Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box

E. Borgonovo¹ , X. Lu¹ , E. Plischke² , O. Rakovec³ , and M. C. Hill⁴ 
¹Department of Decision Sciences, Bocconi University, Milan, Italy, ²Clausthal University of Technology, Clausthal-Zellerfeld, Germany, ³UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany, ⁴University of Kansas, Lawrence, Kansas, USA

Abstract In this work, we investigate methods for gaining greater insight from hydrological model runs conducted for uncertainty quantification and model differentiation. We frame the sensitivity analysis questions in terms of the main purposes of sensitivity analysis: parameter prioritization, trend identification, and interaction quantification. For parameter prioritization, we consider variance-based sensitivity measures, sensitivity indices based on the L^1 -norm, the Kuiper metric, and the sensitivity indices of the DELSA methods. For trend identification, we investigate insights derived from graphing the one-way ANOVA sensitivity functions, the recently introduced CUSUNORO plots, and derivative scatterplots. For interaction quantification, we consider information delivered by variance-based sensitivity indices. We rely on the so-called given-data principle, in which results from a set of model runs are used to perform a defined set of analyses. One avoids using specific designs for each insight, thus controlling the computational burden. The methodology is applied to a hydrological model of a river in Belgium simulated using the well-established Framework for Understanding Structural Errors (FUSE) on five alternative configurations. The findings show that the integration of the chosen methods provides insights unavailable in most other analyses.

1. Introduction

This study focuses on sensitivity analysis of hydrological models, which are used to analyze water shortage (drought), water excess (flood), water quality (contamination of drinking water and/or crops), and river dynamics (erosion). These can cause large socioeconomic damage and ways to prevent such damage are of intense interest. Computer models are developed with the hope of adequately representing the real-world complexity of rainfall-runoff processes in hydrology catchments, the contributing areas from which a given stream/river derives its flow. In this context, adequately suggests a level of accuracy that makes model results useful for managing the system being simulated [Gupta *et al.*, 2012; Foglia *et al.*, 2013]. Model inputs/parameters generally cannot be directly measured in nature with sufficient accuracy, and therefore, are commonly estimated through inverse modeling—see, among others Duan *et al.* [1992] and La Vigna *et al.* [2016]. The process has inherent uncertainty, and measures of uncertainty commonly accompany any modeling analysis [e.g., Pappenberger and Beven, 2006; Montanari, 2007; Beven, 2011; Nearing *et al.*, 2016]. Sensitivity analysis is conducted to understand the relation between inputs and outputs and to obtain insights in what often is a complicated model input-output mapping [Hill and Tiedeman, 2007; Saltelli *et al.*, 2008; Rosero *et al.*, 2010; Mendoza *et al.*, 2015; Norton, 2015; Hill *et al.*, 2016; Pianosi *et al.*, 2016; Razavi and Gupta, 2016; Markstrom *et al.*, 2016; Houle *et al.*, 2017]. These discoveries help the analyst to use simulated values appropriately in planning, risk assessment, and decision support.

In the simulation of environmental/hydrologic systems, the plethora of sensitivity analysis methods that have been developed causes confusion and for many of the methods execution times can be colossal [e.g., Hill *et al.*, 2016, and references cited therein]. All this recommends careful consideration of the needed model runs and how they are used. Of interest are studies that explore how a set of model runs can be used to obtain relevant and varied insights. In this work, we consider the utility of a set of sensitivity analysis methods.

Our approach rests on two main pillars: clearly stating the sensitivity analysis goals from the start and controlling computational burden. Regarding goals, we make use of the methodology of sensitivity analysis settings. A setting is used to frame the sensitivity quest in such a way that the answer can be confidently entrusted

to a well-identified measure [Saltelli et al., 2008, p. 24]. We consider the following three sensitivity settings which have emerged from previous sensitivity analysis studies: *parameter prioritization* [Ratto et al., 2007], *trend identification*, and *interaction quantification*—see among others Borgonovo and Plischke [2016]. These settings can be used at different stages of the modeling process. Saltelli et al. [2000] and Hill and Tiedeman [2007] emphasize the role of sensitivity analysis throughout model development, starting with the model building phase. Ratto et al. [2007] discuss the role of sensitivity analysis to support calibration and validation. (For greater details on the literature on sensitivity analysis in hydrologic modeling, see supporting information section S1.) As an example, consider using the root-mean-square error (RMSE) as the quantity of interest. RMSE measures the distance between the model predictions and the actual physical measurements. In factor prioritization, using RMSE as the quantity of interest produces results that can guide determining which parameters matter the most and least in calibration. Once these are identified, trend identification allows us to understand further whether the dependence of RMSE is monotonic or not on the parameters. Interaction quantification, as specified in Ratto et al. [2007], helps us with identifiability; parameters associated with high individual contribution are more easily identified than parameters owing their importance to interaction effects. (*Small main effect but high total effect: here, such a situation flags an influence mainly through interaction, implying lack of identification* [Ratto et al., 2007, p. 1254]. Or, e.g., high composite scaled sensitivities and large parameter correlation coefficients of Hill and Tiedeman [2007].)

We aim to conduct sensitivity analyses that deliver insights on these sensitivity settings simultaneously while keeping computational burden under control. We propose combining a given-data approach for the estimation of global sensitivity measures [Plischke et al., 2013] with the hybrid local-global method DELSA [Rakovec et al., 2014]. A given-data approach allows us to exploit the data set generated for uncertainty quantification to calculate a variety of global sensitivity measures. We then integrate the insights of global methods with the indications yielded by a method capable of extracting information from the partial derivatives data set.

Our approach addresses each sensitivity setting using multiple sensitivity measures. This is important because any single sensitivity method refers to a particular aspect of the model output response, has theoretical limitations and, moreover, numerical errors might affect the estimates at finite sample sizes. Thus, by relying on an ensemble of sensitivity measures that can be simultaneously estimated, one increases the robustness of the inference without augmenting computational burden. For parameter prioritization, we rely on first-order Sobol' indices, on the δ -importance measure [Borgonovo, 2007], on a sensitivity measure based on the Kuiper metric, a modification of the Kolmogorov-Smirnov distance [Kuiper, 1960; Baucells and Borgonovo, 2013] and on linearized variance index of the DELSA method—see Table 1. For trend identification, we make use of alternative visualization tools for displaying results in an intuitive and easy-to-grasp fashion. Because partial derivatives are the natural sensitivity measures for trend identification, we also use them in this work to create derivative scatterplots (D-scatterplot) jointly with the graphs of the global main effect functions of the functional ANOVA expansion and the Cumulative Sum of Normalized Reordered Output (CUSUNORO) [Plischke, 2010]

plots. Indeed, these last two visualization methods do not require partial derivatives, accommodating the case in which the model execution time required for a partial derivatives data set has not been pursued. Furthermore, we discuss ways to profit from the derivative data set to analyze the regional contribution of the uncertain parameters. There, our goal is to identify whether the importance of a parameter is concentrated in particular ranges of its support (in Ratto et al. [2007], regional sensitivity analysis is associated with a fourth setting, factor mapping). For *interaction quantification*, we use the

Table 1. Summary of the Settings and Sensitivity Methods Used in This Work

Setting/Name	Symbol	Equation
Parameter Prioritization		
First-order Sobol'	η_i, S_i	Equations (2), supporting information equation (S8)
Borgonovo's δ	δ_i	Equation (7)
Kuiper based	β_i^{KU}	Equation (8)
DELSA	S_i^L	Equation (13) ^a
Trend Identification		
Partial derivatives	$\frac{\partial g}{\partial x_i}$	Supporting information equation (S3) ^a
Main effect functions	$g_i(x_i)$	Equation (15)
CUSUNORO	$c_i(u)$	Equation (16)
Interaction Quantification		
Sum of first-order sensitivity indices		
Higher order variance-based indices	$S_{i,j}, ST_{i,j}, \dots$	Supporting information equation (S7)

^aCalculated using local derivatives obtained at distributed points in parameter space.

uncertainty quantification data set and rely on Polynomial Chaos Expansion (PCE) [Sudret, 2008; Marelli and Sudret, 2015], HDMR [Ziehn and Tomlin, 2009], and LASI (a subroutine based on high-dimensional model representations described in supporting information section S3) to provide a direct estimation of the second-order effects from the given sample.

We demonstrate the approach by conducting numerical experiments within the well-established hydrologic modeling Framework for Understanding Structural Errors (FUSE) [Clark *et al.*, 2008]. This framework was the first in the hydrologic sciences to be designed specifically to support consideration of alternative working hypotheses (also called alternative models or multi model analysis) [Clark *et al.*, 2011a]. We provide results for a medium-sized basin situated in the hilly parts of the Belgian Ardennes (Western Europe). We start with a reference configuration (FUSE-016) and then compare it with other four alternative structure configurations while studying the sensitivity of the model RMSE to variations in the parameters. The sensitivity methods allow us to confidently identify the key drivers of RMSE variability across the configurations, to establish whether RMSE is increasing or decreasing in the parameters and to identify the presence of interactions.

The sensitivity analysis insights of this work will be broadly applicable for the next generation modeling frameworks, such as the Structure for Unifying Multiple Modeling Alternatives (SUMMA) [Clark *et al.*, 2015a, 2015b] and the ongoing community-based efforts on parameter regionalization schemes of hydrology/land-surface models [e.g., Mizukami *et al.*, 2017; Samaniego *et al.*, 2017]. They also have considerable utility for climate models and other environmental systems [e.g., Mendoza *et al.*, 2015; Cuntz *et al.*, 2015].

The remainder of the work is organized as follows. Section 2 reviews the sensitivity analysis methods used in correspondence of each setting. Model results are presented and discussed in sections 3 and 4, respectively, and finally conclusions are drawn in section 5. Supporting information provides detailed insights on mathematical formulas and includes graphical results for alternative model structures.

2. Methods I: Review, Definitions, and Properties

This section is organized as follows. In section 2.1, we present relevant notation. Sections 2.2–2.5 present the methods and associate each method with the corresponding sensitivity setting. Finally, in supporting information section S1, we offer a literature review of the use of sensitivity analysis in hydrologic modeling.

2.1. Definitions and Notation

The model output of interest is denoted by y , and the model input-output mapping by a generic multivariate function:

$$y = g(\mathbf{x}), \quad g : \mathbb{R}^k \rightarrow \mathbb{R}, \quad (1)$$

where k is the number of model parameters. Two general situations may be distinguished for practical purposes. The analyst is interested in inspecting the sensitivity of the model output at a point of interest, say \mathbf{x}_0 . In this case, the analyst resorts to a local sensitivity analysis method. Or, the analyst is uncertain about some of the model parameters and is interested in the model response at several locations in the model parameter space. In this case, parameters become random variables. Let \mathbf{X} denote the random model parameter vector. The model output then becomes a random variable Y , related to \mathbf{X} through equation (1), i.e., $Y = g(\mathbf{X})$. Let $\mathcal{X} \in \mathbb{R}^k$ denote the support of \mathbf{X} , i.e., the set of all possible values that the input parameters can assume. As is standard in sensitivity analysis, we denote the joint cumulative distribution function (cdf) of the model parameters by $F_{\mathbf{X}}(\mathbf{x})$, their joint density (pdf) by $f_{\mathbf{X}}(\mathbf{x})$, and the marginal cdf and pdf of X_i as $F_{X_i}(x_i)$ and $f_{X_i}(x_i)$. The cdf and pdf of the model output are denoted by $F_Y(y)$ and $f_Y(y)$, respectively.

2.2. Parameter Prioritization: Sensitivity Measures

In *parameter prioritization*, the analyst's goal is to identify the input parameter that, if fixed to a certain value, would lead to the greatest reduction in the variability of the model output [Saltelli and Tarantola, 2002]. Thus, the degree of statistical dependence between Y and X_i is of concern. The stronger the statistical dependence, the more important we consider the parameter. Global sensitivity measures aim at identifying such dependence over the entire parameter support. In a variance-based sensitivity analysis [Ratto *et al.*, 2007], the intuition is to quantify statistical dependence as the expected reduction in model output variance due to fixing input parameter X_i . Let $\mathbb{V}[Y]$ denote the unconditional variance of Y . Let then $\mathbb{V}[Y|X_i]$

denote the variance of Y given that X_i is fixed. Then, we can define the expected reduction in model output variance as [Homma and Saltelli, 1996]

$$\eta_i = \mathbb{E} \left[\frac{\mathbb{V}(Y) - \mathbb{V}(Y|X_i)}{\mathbb{V}[Y]} \right] = \frac{\mathbb{V}(Y) - \mathbb{E}[\mathbb{V}(Y|X_i)]}{\mathbb{V}[Y]} = \frac{\mathbb{V}[\mathbb{E}[Y|X_i]]}{\mathbb{V}[Y]}. \quad (2)$$

Borgonovo and Plischke [2016] point out that the rationale at the basis of variance-based sensitivity measures is common to other sensitivity measures. Equation (2) can be seen as measuring the separation between the current model output distribution and the conditional model output distribution given X_i through variance reduction. In general, we can consider any operator between $\zeta(P_Y, P_{Y|X_i})$, where P_Y and $P_{Y|X_i}$ are unconditional and conditional distributions (pdf or cdf) of the model output. However, to be a sensible measure of the discrepancy between P_Y and $P_{Y|X_i}$, the operator $\zeta(\cdot, \cdot)$ is required to be null when the two distributions are identical, i.e., $\zeta(P, P) = 0$. Examples of inner operators used in the literature are

$$\zeta^V(P_Y, P_{Y|X_i}) = \frac{\mathbb{V}(Y) - \mathbb{V}(Y|X_i)}{\mathbb{V}[Y]}, \quad (3)$$

for variance-based sensitivity measures [Homma and Saltelli, 1996], the L^1 -norm on density functions,

$$\zeta^{L1}(P_Y, P_{Y|X_i}) = \frac{1}{2} \int_{\mathcal{Y}} |f_Y(y) - f_{Y|X_i}(y)| dy, \quad (4)$$

for the δ -importance measure [Borgonovo, 2007], or the Kolmogorov-Smirnov distance between cdfs in the PAWN method [Pianosi and Wagener, 2015],

$$\zeta^{KS}(P_Y, P_{Y|X_i}) = \sup_{\mathcal{Y}} |F_Y(y) - F_{Y|X_i}(y)|, \quad (5)$$

or the Kuiper distance in Baucells and Borgonovo [2013],

$$\zeta^{KU}(P_Y, P_{Y|X_i}) = \sup_{\mathcal{Y}} (F_Y - F_{Y|X_i}) + \sup_{\mathcal{Y}} (F_{Y|X_i} - F_Y). \quad (6)$$

Baucells and Borgonovo [2013] and Borgonovo et al. [2014] discuss the use of distances between cumulative distribution functions in sensitivity analysis proposing a general definition of which equations (5) and (6) are particular cases. In such work, they underline that some limitations associated with the KS distance can be overcome by its modification in the Kuiper metric [Kuiper, 1960]. In particular, the Kuiper distance puts all percentiles on equal footing [Crnkovic and Drachman, 1996, p. 140]. We therefore use this metric to measure distance between cumulative distribution functions in our work.

Once ζ is defined, we can obtain a global sensitivity measure taking any statistic of the random quantity $\zeta(P_Y, P_{Y|X_i})$. For instance, in the PAWN method [Pianosi and Wagener, 2015], the inner operator is the Kolmogorov-Smirnov distance between cumulative distribution functions and the statistic is its median. To obtain variance-based sensitivity measure, we need to take the expectation of $\zeta(P_Y, P_{Y|X_i})$ over all possible values of X_i . Similarly, taking the expectation of equations (4) and (6), we obtain the well known

$$\delta_i = \frac{1}{2} \mathbb{E}_{X_i} \left[\int_{\mathcal{Y}} |f_Y(y) - f_{Y|X_i}(y)| dy \right], \quad (7)$$

and

$$\beta_i^{KU} = \mathbb{E}_{X_i} \left[\sup_{\mathcal{Y}} (F_Y - F_{Y|X_i}) + \sup_{\mathcal{Y}} (F_{Y|X_i} - F_Y) \right] \quad (8)$$

sensitivity measures.

All global sensitivity measures mentioned so far with the exception of first-order variance-based sensitivity measures possess the nullity-implies-independent property.

2.3. Methods Based on Semilocal and Local Sensitivities

One of the most popular methods based on the randomized evaluation of local sensitivities is the method of Morris [Morris, 1991]. The elementary effect of the i th parameter at \mathbf{x}^0 is defined by

$$EE_i(\mathbf{x}^0) = \frac{g(x_i^0 + \Delta, \mathbf{x}_{-i}^0) - g(\mathbf{x}^0)}{\Delta}, \quad (9)$$

where \mathbf{x}_{-i}^0 is the parameter vector with the exception of x_i , and Δ is a predetermined variation of x_i . Unlike typical local sensitivity calculations, Δ is often large and the run sequence forms a stepwise path instead of a star around \mathbf{x}_{-i}^0 . The calculation of $EE_i(\mathbf{x}^0)$ is averaged over randomized locations \mathcal{X} , forming a winding stairs path. The parameters indicated as least relevant by the average of their elementary effects are then candidates to be fixed at their nominal value [Campolongo *et al.*, 2007]. The recently proposed enhanced version of elementary effects [Cuntz *et al.*, 2015] allows for a more computationally efficient sequential screening.

Another class of sensitivity measures in a parameter fixing setting is the class of derivative-based sensitivity measures. Sobol' and Kucherenko [2009] define

$$v_i^{DGS} = \mathbb{E}_{\mathcal{X}_i} \left[\left(\frac{\partial g}{\partial x_i} \right)^2 \right]. \quad (10)$$

The sensitivity measure in equation (10) is equal to the average of the square of partial derivatives evaluated at randomized locations in the parameter space.

In this work, we employ the DELSA method of Rakovec *et al.* [2014]. DELSA combines methodological properties from three other methods: the method of Morris (MoM) [Morris, 1991], the Sobol' method [Sobol', 2001], and regional sensitivity analysis (RSA) [Hornberger and Spear, 1981]. DELSA uses the local equation for variance estimation V_L [Seber and Wild, 1989, p. 191; Draper and Smith, 1998, pp. 129–130; Aster *et al.*, 2013, p. 224; Lu *et al.*, 2012]:

$$V_L = \left(\frac{\partial g}{\partial \mathbf{x}} \right)^T (\mathbf{X}^T \omega \mathbf{X})^{-1} \left(\frac{\partial g}{\partial \mathbf{x}} \right), \quad (11)$$

which linearly propagates the parameter uncertainty expressed by $(\mathbf{X}^T \omega \mathbf{X})^{-1}$ to obtain the variance of model output—see Rakovec *et al.* [2014, Appendix A] for additional mathematical details. In equation (11), we define \mathbf{X} as a $k \times k$ identity matrix and ω is estimated as the reciprocal variance of the uniform distribution from the parameter prior ranges. Then, the total linearized local variance V_L becomes

$$V_L = \sum_{i=1}^k \left(\frac{\partial g}{\partial x_i} \right)^2 s_i^2, \quad (12)$$

where s_i^2 is the a priori parameter variance of the i th parameter. Finally, the DELSA first-order sensitivity measure of the i th parameter is calculated at each sampling point as

$$S_i^L = \frac{\left(\frac{\partial g}{\partial x_i} \right)^2 s_i^2}{V_L}. \quad (13)$$

Equation (13) is the local fraction of the linearized variance of Y apportioned by X_i . The DELSA indices S_i^L are calculated at randomized locations throughout parameter space. The analyst can then consider the full frequency distribution of these sensitivity measures or any other statistical property for making inference. For instance, the median of the distribution of S_i^L is considered in Rakovec *et al.* [2014] for factor prioritization.

2.4. Trend Identification: Sensitivity Measures

In the *trend identification* setting, we address an essential insight about model behavior, the need to understand whether an increase (decrease) in a parameter leads to an increase (decrease) in the model output. The importance of this setting has been appreciated since the seminal work of Samuelson [1941, p. 97]: *in order for the analysis to be useful it must provide information concerning the way in which our equilibrium quantities will change as a result of changes in the parameters taken as independent data*. As also underlined in Samuelson's work, the appropriate sensitivity measures for this task are signs of partial derivatives. As we shall see, an efficient visualization tool is a derivative scatterplot (D-scatterplot, henceforth). If a derivative data set is not available, we argue that one can make use of the following two methods to still obtain

information on trend identification: visualization of the first-order terms of the functional ANOVA expansion and use of the CUSUNORO plot. Let us start with the former first-order terms.

Let $g(\mathbf{x})$ be integrable and let $F_{\mathbf{x}}(\mathbf{x}) = \prod_{i=1}^k F_{X_i}(x_i)$. Then, g can be decomposed exactly into 2^k components [Efron and Stein, 1981]:

$$g(\mathbf{x}) = g_0 + \sum_{i=1}^k g_i(x_i) + \sum_{1 \leq i < j \leq k} g_{ij}(x_i, x_j) + \cdots + g_{1,\dots,k}(x_1, \dots, x_k), \quad (14)$$

where

$$\begin{aligned} g_0(\mathbf{x}) &= \mathbb{E}_{\mathbf{x}}[g(\mathbf{X})] = \int_{\mathbb{R}^k} g(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \\ g_i(x_i) &= \mathbb{E}_{\mathbf{x}_{-i}}[g(\mathbf{X}_{-i}, x_i) | X_i = x_i] - g_0 = \int_{\mathbb{R}^{k-1}} g(\mathbf{X}_{-i}, x_i) f_{\mathbf{x}_{-i}}(\mathbf{X}_{-i}) d\mathbf{x}_{-i} - g_0 \\ g_{ij}(x_i, x_j) &= \mathbb{E}_{\mathbf{x}_{-ij}}[g(\mathbf{X}_{-ij}, x_i, x_j) | X_i = x_i, X_j = x_j] - g_i(x_i) - g_j(x_j) - g_0 \\ &= \int_{\mathbb{R}^{k-2}} g(\mathbf{X}_{-ij}, x_i, x_j) f_{\mathbf{x}_{-ij}}(\mathbf{X}_{-ij}) d\mathbf{x}_{-ij} - g_i - g_j - g_0 \\ &\dots \\ g_{1,2,\dots,k}(\mathbf{x}) &= g(\mathbf{x}) - \sum_{j=1}^{k-1} \sum_{i_1 < i_2 < \dots < i_j} g_{i_1, \dots, i_j}(x_{i_1}, x_{i_2}, \dots, x_{i_j}) - g_0. \end{aligned} \quad (15)$$

In equation (15), \mathbf{X}_{-i} is a shorthand for $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k)$; g_0 is the expectation of Y ; $g_i(x_i)$ is called the first-order effect function and displays the expected behavior of Y as a function of X_i ; and $g_{ij}(x_i, x_j)$ is the interaction effect of X_i, X_j , etc. The generic effect function $g_{i,j,\dots,r}(x_i, x_j, \dots, x_r)$ has null expectation and two generic effect functions are orthogonal [Sobol', 1993; Li and Rabitz, 2012].

The first-order functions $g_i(x_i)$ in equation (15) can be used to obtain information about sign of change. By definition, $g_i(x_i)$ is the conditional expectation of Y given $X_i = x_i$. Thus, $g_i(x_i)$ conveys the average behavior of Y as a function of x_i . Moreover, the first-order effect function $g_i(x_i)$ retains the monotonicity of the original input-output mapping [Beccacece and Borgonovo, 2011]. That is, if $g(\mathbf{x})$ is increasing, then all the $g_i(x_i)$'s are increasing. Then, the visualization of the graphs of the first-order effect functions provides an indication about the expected trend of Y as a function of X_i .

The CUSUNORO curve for parameter X_i is given for $u \in [0, 1]$ by

$$c_i(u) = \frac{u}{\sqrt{\mathbb{V}[Y]}} \mathbb{E}[Y - \mathbb{E}[Y] | X_i \leq F_{X_i}^{-1}(u)] = \frac{1}{\sqrt{\mathbb{V}[Y]}} \int_{-\infty}^{F_{X_i}^{-1}(u)} \mathbb{E}[Y - \mathbb{E}[Y] | X_i = x] dx. \quad (16)$$

By construction, it is a curve with $c_i(0)=0$ and $c_i(1)=0$. The curve $c_i(u)$ displays the average mean of the standardized output when the associated parameter is less than a given quantile u . We may therefore speak of a partial mean to the left (given by $F_{X_i}(x) \mathbb{E}[Y - \mathbb{E}[Y] | X_i \leq x]$) in contrast to the conditional mean to the left given by $\mathbb{E}[Y - \mathbb{E}[Y] | X_i \leq x]$ and note that due to standardization, this mean to the left and the corresponding mean to the right add up to zero. If the model is an increasing function of X_i , then the partial mean to the left is always lagging behind the global mean. Therefore, the CUSUNORO curve is negative for all values of u . Conversely, if the model is decreasing in X_i , then the CUSUNORO curve is positive for all values of u . It can also be proven that if there exists a linear regression curve with respect to the rank of the parameters:

$$\mathbb{E}[Y - \mathbb{E}[Y] | \hat{F}_X^{-1}(X) = u] = \alpha(u - \frac{1}{2}), \quad (17)$$

the CUSUNORO curve has a local extremum at $u_0 = \frac{1}{2}$. Hence, any extreme value not located in the center of the CUSUNORO plot shows a nonlinear dependence between Y and X_i .

2.5. Interaction Quantification: Sensitivity Measures

In an interaction quantification setting, we are interested in understanding whether the model response is additive or not. If the response is additive, then the variation of the output is the direct sum of the individual

effects of the variations in the parameters. Herein, we aim at studying interactions while remaining in a given-data frame. As underlined in *Saltelli et al.* [2000] the quantity

$$S^{\text{Interaction}} = 1 - \sum_{i=1}^k S_i \quad (18)$$

can be considered as an indicator of the percentage of the model output variation apportioned by interactions. Because first-order variance-based sensitivity indices can be estimated from a given-data frame, this quantity $S^{\text{Interaction}}$ is also delivered by a given-data approach. Then, if the sum of S_i (or η_i) is close to 1, we are informed that interactions provide a limited contribution to the model output variation, so that the model response can be regarded as additive. Conversely, further investigation on the nature of interactions is needed. Several methods are available. For instance, one can start investigating the effects of the interactions of all pairs, through linear inferential measures [Hill and Tiedeman, 2007; Hill et al., 2016]. Herein, we rely on second Sobol' sensitivity indices $S_{i,j}$ —see supporting information section S3 for the mathematical definitions. Alternative ways are available for estimating second-order Sobol' sensitivity indices directly from the uncertainty quantification sample. We employ here two direct methods based on the high-dimensional model representation (HDMR) theory (see also supporting information section S3) and we compare them to a brute force estimation method that makes use of a Kriging emulator as an intermediate step. We refer to supporting information section S4 for technical details about estimation cost. The next section discusses results and insights for a hydrological case study.

3. Application

3.1. Hydrological Framework

The ensemble of sensitivity methods described in the aforementioned sections is executed using a set of models developed to simulate a medium-sized catchment (Lasnenville, 200 km²) located in the Belgian Ardennes (Western Europe). The maritime climate can be classified as rain dominated with irregular snow in winter. The runoff regime is highly variable with low summer discharges and high winter discharges. The annual precipitation yields around 1000 mm and mean annual air temperature is 7.5°C. Mixed-forest and agricultural areas represent the two dominant land cover classes [Rakovec et al., 2012].

Five models are developed using the Framework for Understanding Structural Errors (FUSE), a well-established modular framework, which enables constructing a suite of hydrological models to rigorously implement and evaluate hydrological theories [Clark et al., 2008, 2011b]. The ability of a model to adequately approximate dominant hydrological processes depends on (1) the choice of state variable in the unsaturated and saturated zones and (2) the choice of flux equations describing the surface runoff, vertical drainage between soil layers, base flow, and evapotranspiration [Clark et al., 2008].

The vertical dimension of all models is discretized into two reservoirs: the unsaturated reservoir above the water table (often referred as soil moisture storage) and the saturated one below the water table (also known as groundwater storage). The outflow of the two model reservoirs constitutes the total simulated river flow, which is often also called river discharge and/or streamflow. Model output g used to assess model sensitivity of this study is an aggregated metrics over T daily time steps. The metric g is defined to quantify discrepancies between the model and reality (real-world measurements/observations): the root-mean-square error (RMSE) between the simulated streamflow (q_{sim}) and the observed streamflow (q_{obs}):

$$g = \text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (q_{obs,t} - q_{sim,t})^2}. \quad (19)$$

The FUSE-016 configuration has a “single-layer” architecture for the unsaturated zone, which does not allow for vertical variability in soil moisture. Evapotranspiration is restricted to the upper unsaturated zone and is a linear function of storage between wilting point and field capacity. The FUSE-016 does not allow any vertical drainage when saturation is below field capacity. FUSE-016 has a single nonlinear groundwater reservoir of unlimited size. The surface runoff is conceptualized using the “ARNO/VIC” parameterization, and the routing schemes employ the time delay function using a gamma distribution. The FUSE-014 and FUSE-160 models extend the FUSE-016 configuration by alternative evapotranspiration processes from the unsaturated

Table 2. The Ranges for Parameters of Five FUSE Models are as Defined in *Rakovec et al.* [2014], Except for the Lower Bound of TIMEDELAY^a

No.	Parameter Name	Description	Units	Lower Limit	Upper Limit
1	MAXWATR_1	Maximum storage in the unsaturated zone	mm	50	500
2	MAXWATR_2	Maximum storage in the saturated zone	mm	25	250
3	FRACTEN	Fraction total storage as tension storage		0.05	0.95
4	PERCRTE	Vertical drainage rate	mm/d	0.01	1000
5	PERCEXP	Vertical drainage exponent		1	20
6	BASERTE	Base flow depletion rate	mm/d	0.001	1000
7	QB_POWR	Base flow exponent		1	10
8	AXV_BEXP	ARNO/VIC "b" exponent for surface runoff		0.001	3
9	LOGLAMB	Mean of the log-transformed TI ^b distribution	m	5	10
10	TISHAPE	Shape parameter for TI ^b distribution		2	5
11	TIMEDELAY	Routing parameter (time delay in runoff)	Day	0.1	2
12	FRCHZNE	Fraction of tension storage in the primary zone (unsaturated zone)		0.05	0.95
13	FPRIMQB	Fraction of free storage in the primary reservoir (saturated zone)		0.05	0.95
14	RTFRAC1	Fraction of roots in the upper soil layer		0.05	0.95
15	PERCFRAC	Fraction of drainage to tension storage in the lower layer		0.05	0.95
16	FRACLOWZ	Fraction of soil excess to lower zone		0.05	0.95
17	QBRATE_2A	Base flow depletion rate for the primary reservoir	Day ⁻¹	0.001	0.25
18	QBRATE_2B	Base flow depletion rate for the secondary reservoir	Day ⁻¹	0.001	0.25

^aSmaller values of TIMEDELAY produce unreasonable results given the 1 day time step of the model. Note that the parameters 1–11 belong to the FUSE-016 configuration. Parameters 12–18 belong to extra processes incorporated within alternative model structures of FUSE-014, FUSE-160, FUSE-072, and FUSE-170.

^bTI: topographic index.

zone, which is represented by two cascading reservoirs. The FUSE-072 model enables for vertical drainage through a nonlinear function, which is the only difference with respect to FUSE-016. The FUSE-170 configuration addresses alternative representations of the base flow parameterizations with respect to FUSE-016 by employing two linear groundwater storages. We refer to *Clark et al.* [2011b] for greater details.

The number of parameters for the five FUSE models ranges between 11 and 14 [*Clark et al.*, 2008; *Rakovec et al.*, 2014]. The parameters are summarized in Table 2. The parameters cannot be directly measured in nature with sufficient accuracy, and are location specific based on the regional climate and physiographic basin properties.

This study makes use of the model simulations at daily time step presented by *Rakovec et al.* [2014] for a 10 year period from 1 October 1998 to 30 September 2008. The parameter ranges applied in this study are slightly adjusted from *Clark et al.* [2011b] based on our prior knowledge. Note that the sample size N of this study is 9548, which represents the number of base model runs for the FUSE-016 model. The difference from the 10,000 runs used in *Rakovec et al.* [2014] originates from revising the lower parameter bound for TIMEDELAY from 0.01 to 0.1.

In the remainder of this section, we focus the presentation on results for the FUSE-016 configuration, while the four alternative models FUSE-014, FUSE-160, FUSE-072, and FUSE-170 are used to assess the robustness of parameter sensitivity analysis methods for alternative model structures. Results for these models are described in the supporting information, namely Figures S3–S5.

3.2. Parameter Prioritization: Results

For identifying the most important parameters, we use an ensemble of sensitivity indices combining indications from variance-based, density-based and cdf-based global sensitivity measures. Specifically, using the given-data estimators described in the supporting information section S2, we estimate first-order Sobol' indices η_i , Borgonovo's δ , and the Kuiper index β^{KU} .

Figure 1 shows the boxplots of the bootstrap estimates for these three global sensitivity measures, with a bootstrap sample size $B = 500$. For a discussion about computational cost, see supporting information section S2.3 and supporting information Figure S1. All three approaches rank TIMEDELAY, AXV_BEXP, and FRACTEN as the most influential parameters. The first two parameters directly influence the dynamics of simulated streamflow, in particular its timing and magnitude (TIMEDELAY), and the partitioning of incoming precipitation into quickly responding surface runoff and slow base flow components (AXV_BEXP). Their role, therefore, explains the direct and strong influence on the RMSE in equation (19), which is derived directly

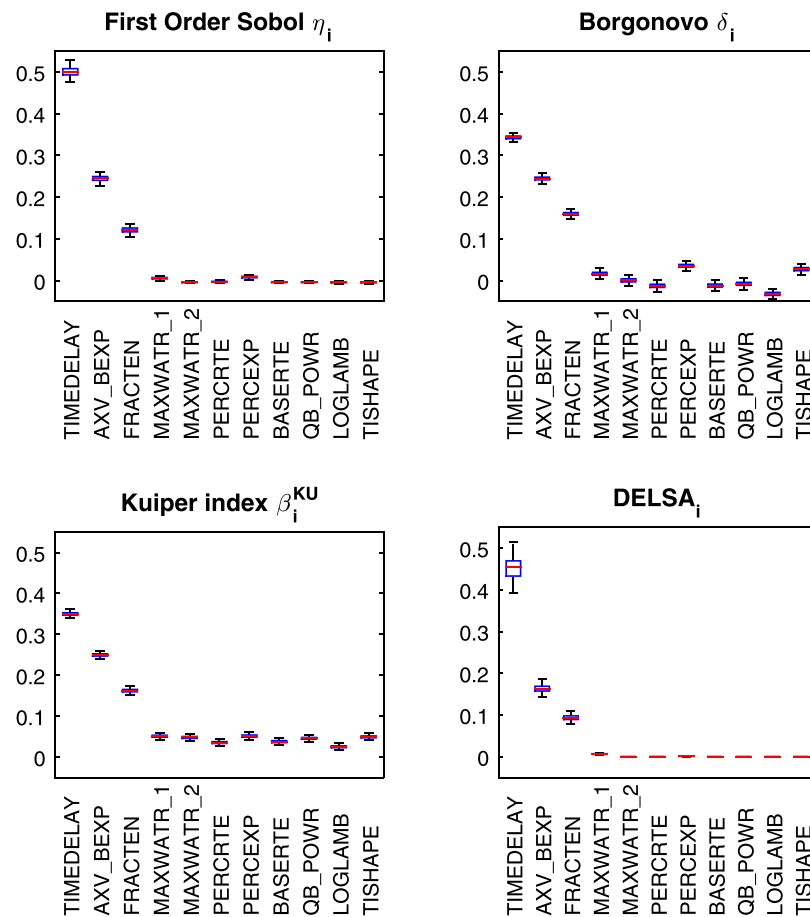


Figure 1. Factor prioritization using four methods. Boxplots for the bootstrap estimates for three global sensitivity measures (η_i , δ_i , β_i^{KU}), with 500 bootstrap replicates ($C = 9548$). Similarly, the bootstrap median of DELSA is presented in the fourth graph. All sensitivity indices agree in suggesting TIMEDELAY, AXV_BEXP, and FRACTEN as key uncertainty drivers. The small bootstrap error bars show that the sensitivity measures are confidently assessed at the available sample size.

from the simulated streamflow. The third most influential parameter (FRACTEN) has a direct effect on the soil moisture dynamics. It quantifies tension storage as a nonlinear function of the total storage in the unsaturated zone. FRACTEN closely controls the magnitude of evapotranspiration processes, i.e., return of incoming precipitation back to the atmosphere, and it also indirectly affects the magnitude of total modeled streamflow. Overall, the importance of the three key parameters is identified clearly and consistently, which is shown by the narrow and not overlapping bootstrap uncertainty bounds.

Furthermore, Figure 1 includes the parameter ranking obtained using the DELSA index S_i^L , which is based on the median cdf value. The results are in clear agreement with the results produced by the three global sensitivity measures estimated directly on the uncertainty quantification sample.

We provide a visual complement to these results in supporting information section S2.4 and supporting information Figure S2.

To corroborate these results, we use the median statistic of the DELSA in equation (13) (Figure 1). This graph identifies LOGLAMB, PERCRTE, and BASERTE as least relevant parameters, which is consistent with the ranking of global sensitivity measures. Thus, for the RMSE of FUSE-016, the key uncertainty drivers and the parameters to fix at their base case are clearly identified.

Figure 2 displays the spatially distributed derivative-based sensitivities, with the goal of identifying important parameter regions, in a factor mapping setting. For illustration purposes, we limit our attention to four parameters (columns in Figure 2). Figure 2 row (a) presents graphs with the S_i^L sensitivities on the horizontal axis and the model output value on the y axis, as per Rakovec et al. [2014]. These results show a crucial

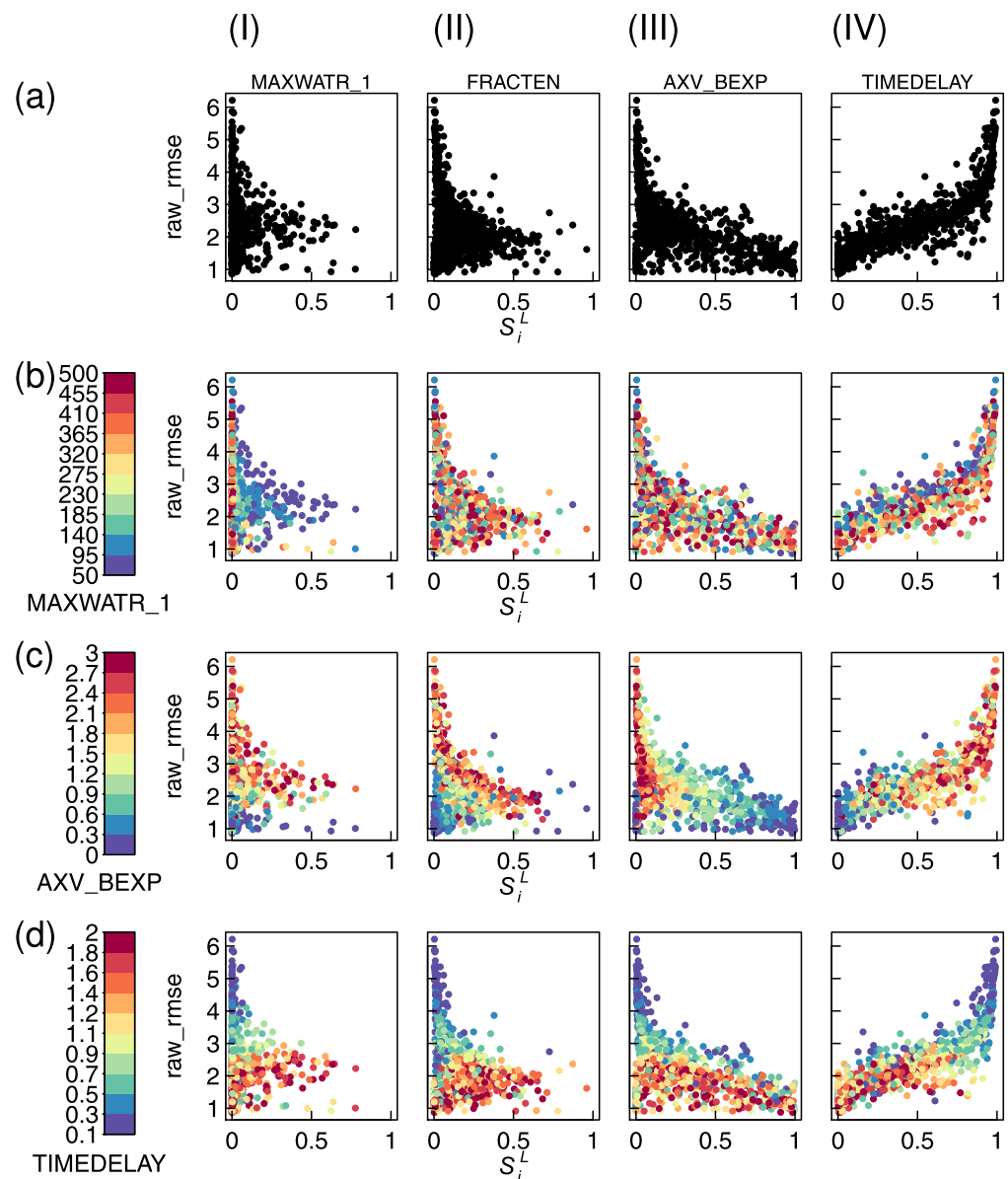


Figure 2. Prioritization using DELSA. DELSA results showing parameter importance, measured using first-order metric S_i^L , plotted with the model output root-mean-squared error (RMSE). Each dot represents scaled local sensitivities calculated for one set of parameter values. Nine thousand five hundred forty-eight dots are shown in each figure. The RMSE for each dot is the same in each figure; the S_i^L value changes. (a) Black and white figures emphasize the position of the dots for the parameters in columns (I–IV). (b–d) The dots are colored based on the value of the parameter listed below the color bar.

difference in the contribution of TIMEDELAY, and AXV_BEXP, the parameters associated with the largest sensitivities. TIMEDELAY is most important in the subset of parameter values associated with higher RMSE, i.e., where the performance of the model is poorer. Conversely, AXV_BEXP is important in regions of lower RMSE, i.e., where the model has a better prediction capability. These considerations show that using information coming from derivatives in a regionalized DELSA setting allows the analyst to delve deeper into how each parameter contributes to the RMSE variability. This leads to insights that enrich and complement the information in Figure 1.

For this model, the partitioning of incoming precipitation into fast and slow flow components governed by AXV_BEXP may be more important than the routing dynamics characterized by the TIMEDELAY parameters, if the focus was on the “acceptable” model performance-simulations. Additionally, although FRACTEN and

MAXWATR_1 exhibit considerably less pronounced importance, some parameter combinations yield $S_i^L > 0.5$, and some of these are very good fitting models based on RMSE.

Figures 2b–2d yield additional insights. *Random color scatter* indicates that the value of parameter importance (measured here using S_i^L) and model fit (measured here by RMSE) are unrelated to the value of the parameter (rows b–d). Thus, the value of the MAXWATR_1 parameter (row b) is mostly inconsequential to the results shown for FRACTEN and AXV_BEXP (columns II and III). The only possible pattern is that the worst-fitting models appear to be dominated by small values of MAXWATR_1.

If the dot color changes vertically in the plots, then model fit depends on the parameter value. For instance, Figure 2, plot IVb shows that for any parameter importance level (for parameter TIMEDELAY), poorer fitting models are dominated by larger values of the MAXWATR_1 parameter. Figure 2 plot Ic, and all of row d show vertical patterns in dot color.

If the dot color changes horizontally in the plots, then parameter importance depends on the value of the parameter. For example, Figure 2 plot Ib shows that nearly all models with large values of MAXWATR_1 are insensitive to the MAXWATR_1 parameter. Figure 2 plot IVd shows that large sensitivities for the TIMEDELAY parameter (the most important parameter) are related to values of the TIMEDELAY parameter smaller than about 1 day. The time step of the model used is 1 day, and this suggested that evaluation of models with these small TIMEDELAY values are worth considering closely.

3.3. Trend Identification: Results

For trend identification, we consider two situations: in the first case, the available data set comprises partial derivatives; in the second case, the analyst has available only the sample of input-output realizations.

In the first case, the sign of the partial derivatives immediately identifies direction of change. Figure 3 shows the D-scatterplot for the FUSE-016 model, plotting the derivatives made available by the DELSA method. Light color dots (red) refer to positive values, darker dots (blue) to negative values. Figure 3 (top left) shows that the derivatives of RMSE with respect to TIMEDELAY estimated at several locations are negative. Thus, we expect that an increase in TIMEDELAY has a decreasing effect on the RMSE. Figure 3 (top middle) plots the derivatives of RMSE with respect to AXV_BEXP. We observe both positive and negative values, which implies that an increase in AXV_BEXP does not necessarily lead to an RMSE increase. Specifically, if we look at region $[0, 1]$, we observe both light (red) and dark (blue) dots in the graph (Figure 3 (top middle)). For values of AXV_BEXP greater than unity, however, we observe mainly light dots, which indicates that the effect of an increase in AXV_BEXP leads to an increase in RMSE. Figure 3 (top right) presents the derivatives of RMSE with respect to the parameter FRACTEN. Again, the existence of both positive and negative values implies nonmonotonicity. However, we observe a majority of positive derivatives (light dots (red)), which indicates an over-all positive effect. The remaining plots of Figure 3 show the existence of both positive and negative derivatives for the remaining parameters.

Insights on direction of change can also be directly obtained from the original data set. A first way is to plot the first-order effect functions in Figure 4, where we present the COSI curves (in red) together with the input-output scatterplots. One can observe that TIMEDELAY shows a nonlinear decreasing effect on the RMSE, while, in general, AXV_BEXP and FRACTEN present an ascending trend. In particular, AXV_BEXP shows wiggling trend on its support $[0, 1/2]$, which is consistent with result of D-scatterplot, where nonmonotonic effect is implied in the same region. Besides, PERCEXP shows slightly increasing effect. For the remaining parameters, there is no strong evidence of decreasing or increasing first-order effect.

Figure 5 illustrates the CUSUNORO plot of FUSE-016 data. Each curve refers to a given parameter. Curves above the horizontal zero line signal a decreasing effect, curves below the horizontal zero line suggest the opposite. Parameter TIMEDELAY (circle curve) is associated with the CUSUNORO curve that shows the highest peak above the zero horizontal axis and is therefore the parameter with the strongest negative impact on the RMSE. This is in accordance with the information provided by the first graph in both Figures 3 and 4. Similarly, parameter MAXWATR_1 (dashed curve) has a negative effect. Conversely, parameters AXV_BEXP (triangle curve) and FRACTEN (square curve) have an increasing effect on the RMSE. Besides, the magnitudes of the deviations from the zero horizontal line can be used to infer information about the strength of impact. Figure 5 indicates TIMEDELAY, AXV_BEXP, and FRACTEN as the three most relevant parameters, in accordance with previous findings. Furthermore, the vertical asymmetry of the CUSUNORO curve implies

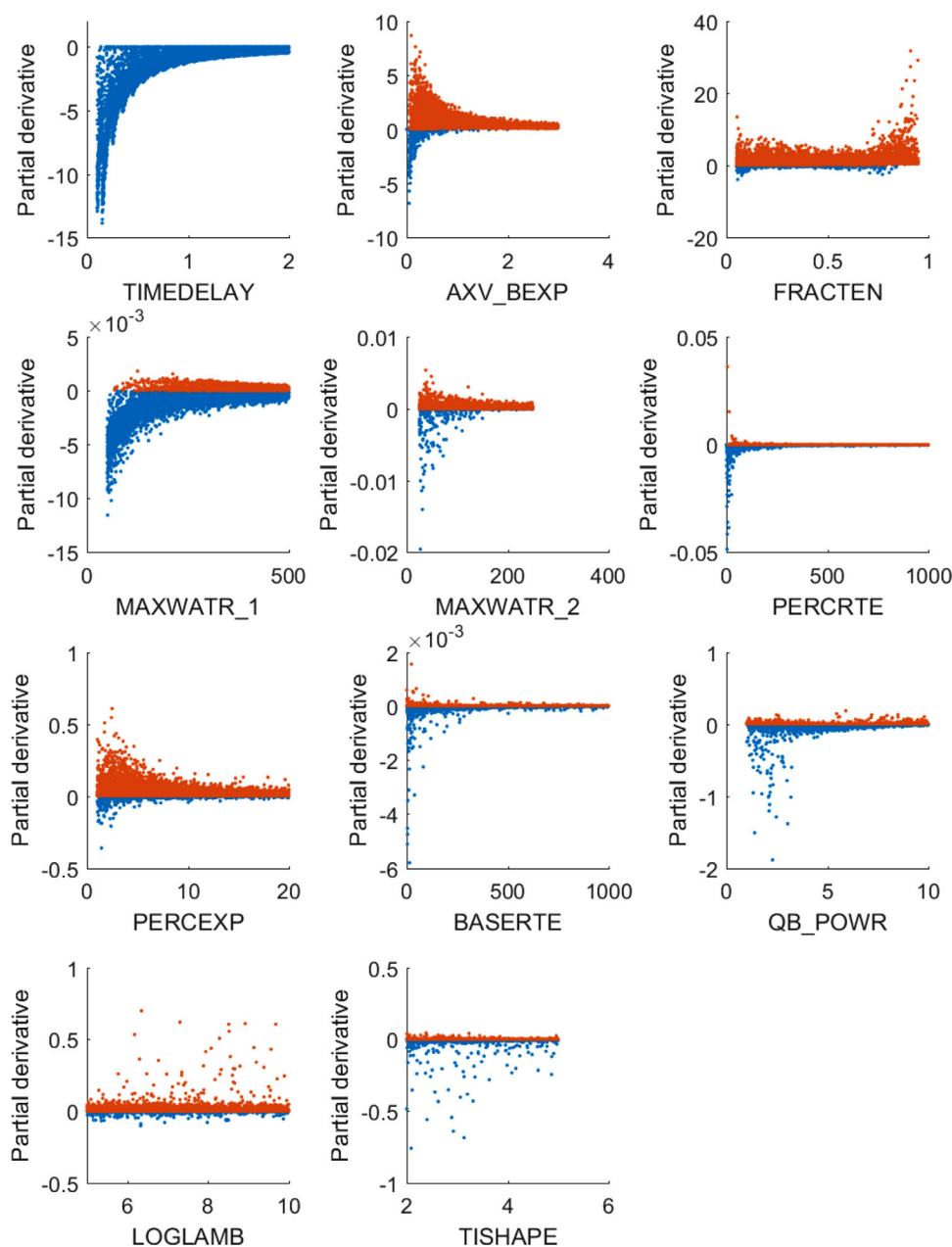


Figure 3. Trend identification using D-scatterplot for the eleven parameters of the FUSE-016 model. The vertical axis in each plot displays the derivative of the RMSE with respect to the corresponding parameter. Note: the axis is not normalized, because the goal of the plot is to indicate sign (trend) and not importance.

the nonlinearity of the first-order effect. For instance, TIMEDELAY and AXV_BEXP are slightly asymmetric to the right (have steeper left parts), which implies that we can expect a nonlinear first-order effects. This result is, again, consistent with the graphs of the first-order effect functions in Figure 4.

3.4. Interaction Quantification: Results

A well-established method for the identification of interactions is to check the sum of first-order Sobol' indices. From Figure 1, we observe that, the sum of first-order Sobol' indices is about 90%, indicating that interactions have limited relevance. Thus, the RMSE can be considered a nearly additive function of the parameters over the ranges of interest.

However, to investigate further, we study second-order interactions, calculating the second-order Sobol' indices S_{ij} . We compare three different estimation methods: PCE is implemented in UQ_lab [Sudret, 2008;

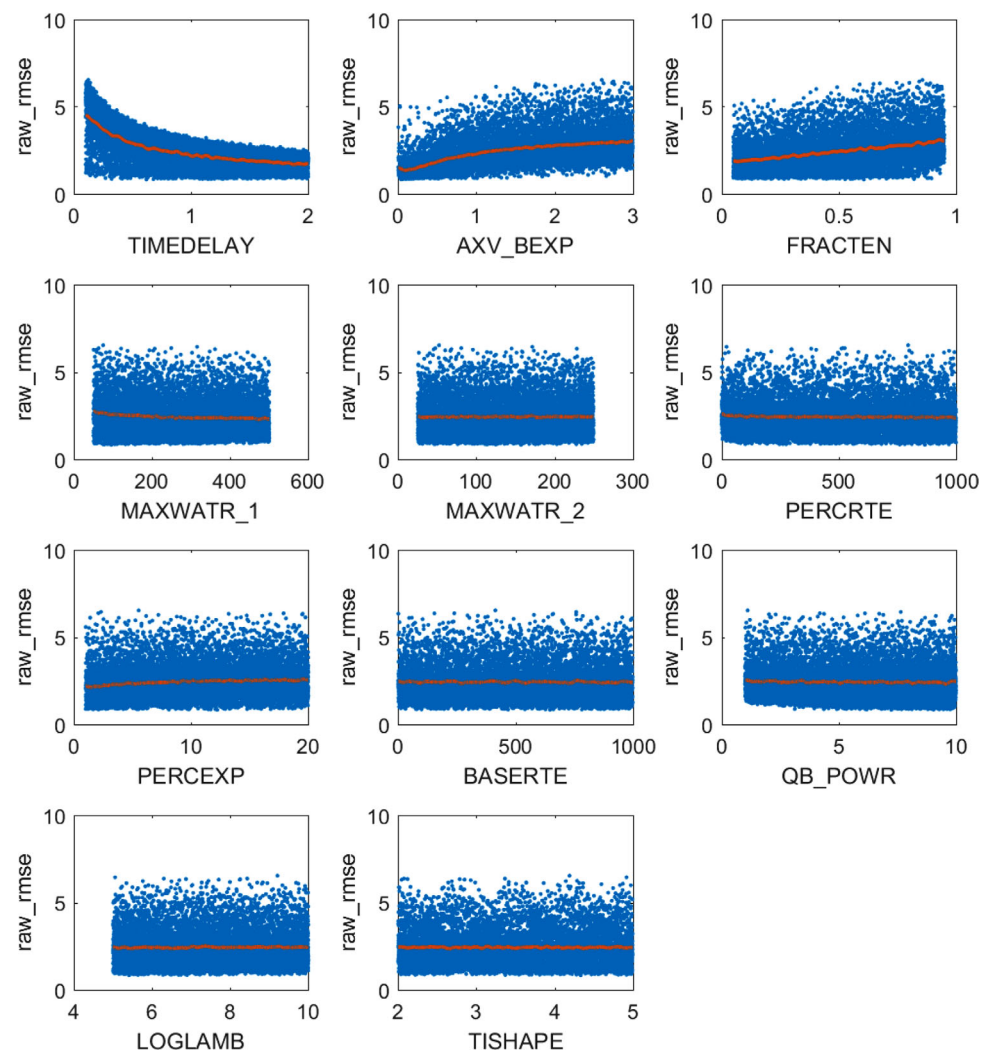


Figure 4. Trend identification using first-order effect $g_i(x_i)$ for the FUSE-016 model. COSI subroutines (red lines) and input-output scatter-plots (blue dots) are shown.

Marelli and Sudret, 2015], HDMR [Ziehn and Tomlin, 2009], and LASI (see supporting information section S3). The HDMR and LASI subroutines allow to estimate the second-order indices directly from the available 9548 model input-output realizations. The PCE subroutine is trained on a subsample of size 2000 (this is the largest at which calculations can be performed on the available pc without encountering an out-of-memory error). Because we have 11 model parameters, we have 55 second-order interactions. Figure 6a illustrates the second-order interactions associated with the five largest values of S_{ij} .

The three most important parameters, TIMEDELAY, AXV_BEXP, and FRACTEN, are involved in the most relevant second-order interactions. In particular, the interaction between TIMEDELAY and AXV_BEXP is identified as the strongest second-order interaction. However, we need to observe that the values of second-order Sobol' indices are small. The last two plots of Figure 6a (estimated via HDMR and LASI subroutines) report that all estimates of S_{ij} are less than 0.05. This confirms the observation stated at the beginning of this section, that interactions have limited influence in the determination of the RMSE for the configuration of interest.

To investigate further and interpret these results in terms of identifiability, we report results for the total-order indices. According to Ratto et al. [2007], Saltelli et al. [2008], and others in fact, parameters associated with a small main effect but having a high total effect evidence a lack of identifiability. To offer an intuitive visual interpretation, we refer to Figure 6b. Here the value of first-order Sobol' index for each model input is on the horizontal axis, while the vertical axis presents the value of total-order Sobol' indices. Each dot in the

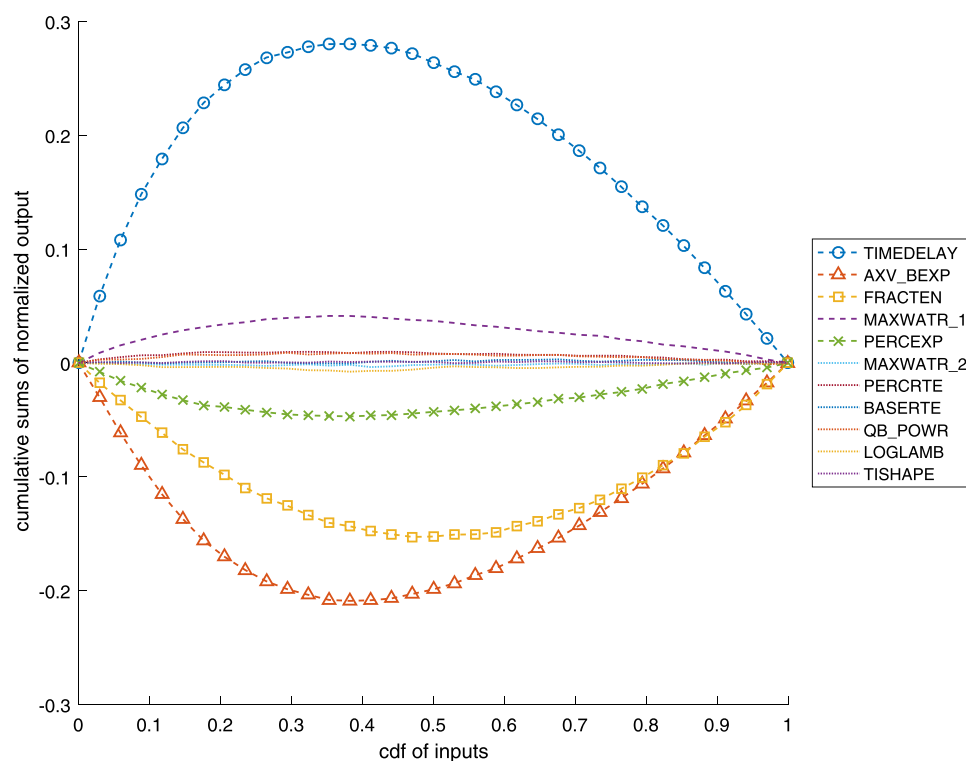


Figure 5. Trend identification using the CUSUNORO plot for FUSE-016 model. Curves above the zero horizontal line indicate a decreasing effect (model output decreases with increasing parameter values). Curves below the horizontal axis show an increasing effect (output increases with increasing parameter values). Curves aligned with the zero horizontal line show a negligible effect.

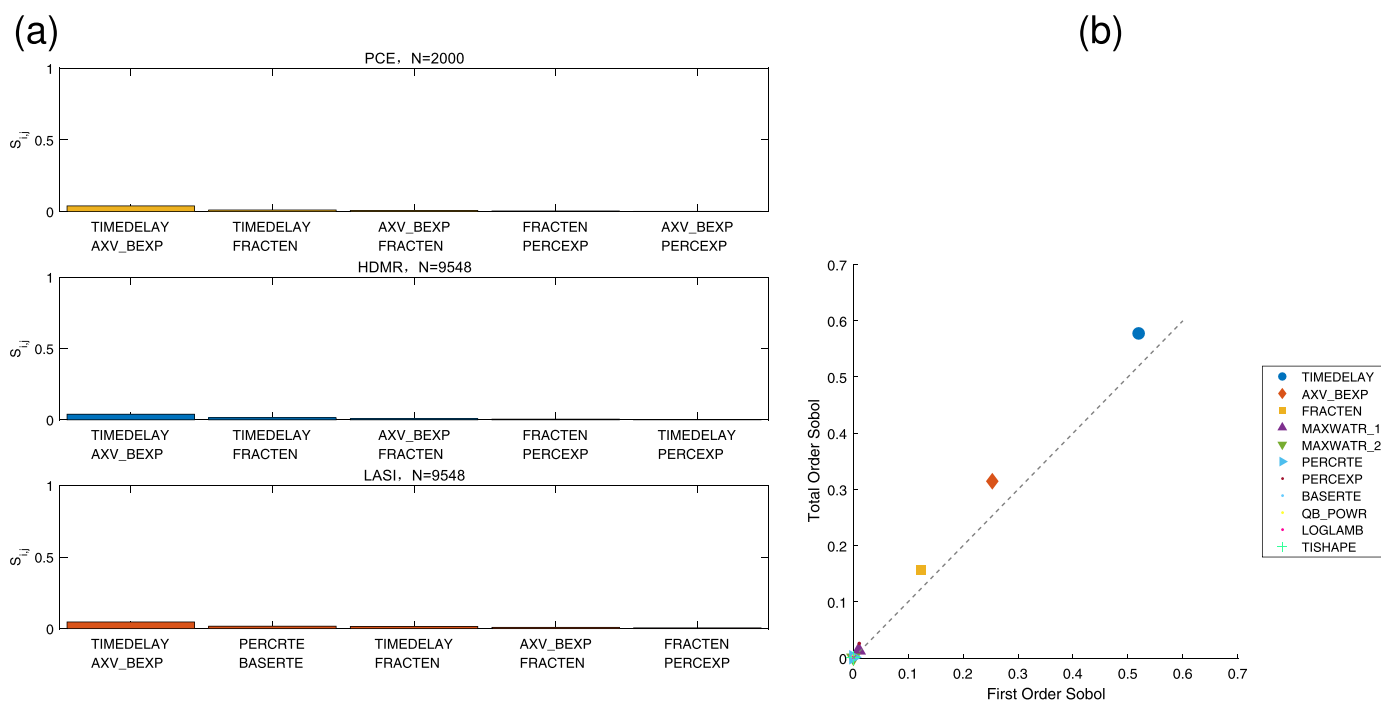


Figure 6. (a) Interaction quantification using the five highest estimates of second-order Sobol' indices S_{ij} for FUSE-016 model. Calculated using subroutines PCE, HDMR, and LASI. All methods register the value of S_{ij} for $i = \text{TIMEDELAY}$ and $j = \text{AXV_BEXP}$ as highest second-order index. However, the values of second-order Sobol' indices are small (the indices may run on a scale between 0 and unity), confirming that interactions have a limited effect on RMSE in this case. (b) First-order (S_i) plotted against the corresponding total-order (S_T) Sobol' indices for the 11 FUSE-016 parameters. Estimation is performed using the PCE subroutine in UQ_Lab from the available uncertainty quantification sample. Values on the line indicate no parameter interaction. Hill et al. [2016] show an example of this graph for parameters with greater interaction.

graph corresponds to one parameter. Less identifiable parameters would lie further toward the top-left corner. For the FUSE-016 configuration, all parameters lie close to the 45° line, confirming the limited contribution of interaction effects. Thus, we register very limited identifiability issues for this configuration.

3.5. Alternative Model Configurations: Results

While in the description of results we have focused on the FUSE-016 configuration, we performed similar sensitivity analyses for the FUSE-014, FUSE-072, FUSE-160, and FUSE-170 models described by Rakovec *et al.* [2014]. These results are summarized in the supporting information accompanying this paper (supporting information Figures S3–S5). We describe here the main differences in the sensitivity analysis results with respect to the FUSE-016 configuration, discussing them setting by setting. Regarding parameter prioritization, supporting information Figure S3 shows that across all FUSE configurations the two most important parameters are TIMEDELAY and AXV_BEXP. The third most important parameter is PERCEXP for FUSE-014, FUSE-072, and FUSE-170 model. The parameter FRACTEN, the third most important for the FUSE-016 configuration, ranks fifth for FUSE-014 and fourth for FUSE-014, FUSE-072, and FUSE-170. For FUSE-160, FRACTEN is ranked third by first-order Sobol' sensitivity measures and median DELSA indices, while it ranks fourth based on the moment-independent sensitivity indices δ and β^{ku} . The parameter RTFRAC1 ranks third based on these sensitivity measures and fourth based on first-order Sobol' sensitivity measures and median DELSA. This parameter is not present in the FUSE-016 configuration. For trend identification, we obtained the D-scatterplot, the COSI and CUSUNORO plots for all configurations. The CUSUNORO results are presented in the plots of supporting information Figure S4. The direction of change associated with the most important parameters remains the same across all the different configurations. Specifically on FUSE-160, the CUSUNORO curve of parameter RTFRAC1 shows a prevailing decreasing trend on the RMSE—see supporting information Figure S4b. To test this indication further, consider the corresponding D-scatterplot (supporting information Figure S5). The first plot in supporting information Figure S5 shows that RTFRAC1 has a nonmonotonic effect, however, with a prevalence of negative partial derivatives. Positive values of the partial derivatives are only associated with high values of this parameter. Overall, the D-scatterplots of supporting information Figure S5 confirm the observations of the other configurations, though indicating that also in the FUSE-160 configuration the RMSE is not a monotonic function of the parameters.

Finally, for interaction quantification, we computed the second-order and total-order Sobol' indices measures using the PCE, HDMR and LASI subroutines. Results show a negligible effect of interactions in all configurations, with the sum of all second-order indices never exceeding 6% of the total variance. This value indicates that identifiability issues are low and that the five model configurations are not overparameterized, as suggested by Rosero *et al.* [2010].

4. Discussion

The investigation we have carried out leads to four takeaways for the analysis of hydrological models as well as for the practice of sensitivity analysis:

1. Given the huge number of available methods, the search of the appropriate sensitivity method needs to be made systematic. The most rigorous way to make the analysis systematic is through the formulation of a sensitivity analysis setting. A setting allows the analyst to transparently choose the method that answers the sensitivity analysis question at hand.
2. Employing a unique sensitivity method for performing a sensitivity analysis even within the same setting is suboptimal. In fact, each sensitivity analysis method has merits as well as limitations. To illustrate, take an analysis performed within a trend identification setting. Here derivatives suggest the direction of change of the model output as a result of changes in the model parameters. However, they require scaling to suggest the importance of a parameter, especially if parameters have different units—in that case, unscaled partial derivatives are not even comparable. An analyst willing to identify the most important parameters needs to take sensitivity measures in a factor prioritization setting. Here a desirable property is that nullity of the sensitivity measure implies independence. First-order variance-based sensitivity measures, while appropriate, do not possess this property. Thus, complementing the analysis with the calculation of a moment-independent method increases the robustness of the inference. If the ranking of variance-based methods is confirmed by the ranking of a moment-independent method, we gain additional confidence about what parameter is important, without the need of additional model runs.

3. Alternative methods may be applied under different circumstances. Consider again a trend identification setting. Partial derivatives are available through the DELSA method. However, in case derivatives are not available, one can use a CUSUNORO plot joint with the plot of the first-order effects of the functional ANOVA expansion to still obtain insights about trend.
4. Each sensitivity method was supposed to require a specific design, calculating several sensitivity measures simultaneously would become a nightmare. However, we can now estimate efficiently several sensitivity measures simultaneously from the same model output sample. This then allows to exploit the synergies and complementary insights that such sensitivity measures make available to the modeler.

We believe that our study contributes to building a comprehensive sensitivity analysis framework which enables a thorough characterization relevant sensitivity-related properties of model responses, as recently advocated by Razavi and Gupta [2015]. Our approach can be extended to more complex environmental models currently being developed in the hydrological community, such as the next generation modeling framework SUMMA [Clark *et al.*, 2015a, 2015b]. SUMMA considers water and energy closure together and also allows to fully solve Richards' equation of the unsaturated flow. All these factors yield much higher degrees of input parameter uncertainty than in this study and the proposed framework can be directly employed for uncertainty quantification and sensitivity analysis of this new generation of hydrological models.

Finally, while some aspects of our analysis are specific for hydrological modeling, the approach is applicable to the statistic diagnosis of models used in the broader environmental and climate literature. For example, there is a growing interest in the sensitivity analysis of integrated assessment models for climate change. Works such as Confalonieri *et al.* [2010], Anderson *et al.* [2014], Butler *et al.* [2014a, 2014b], Gao *et al.* [2016], Marangoni *et al.* [2017], and Paleari and Confalonieri [2016] show the growing trend of sensitivity analysis investigations in the climate and environmental modeling arena, where also the need for keeping computational burden under control is felt. In this respect, the ensemble approach developed here might result in supporting investigators in such sectors as well.

5. Conclusion

In the simulation of environmental/hydrologic systems and climate models, sensitivity analysis methods have the potential of yielding crucial insights that allows analysts to fully exploit modeling efforts. Our investigation introduces an approach that (1) makes the derivation of insights systematic and (2) controls computational burden. We have explored insights concerning parameter prioritization, trend identification, and interaction quantification. All these insights provide the analyst with a deeper understanding of the hydrologic model at hand and, in turn, of the relevant phenomena. The present approach is especially promising given the next generation of hydrological models, the SUMMA, in which several problems associated with the FUSE modeling framework are overcome from the modeling side.

Notation

ANOVA	analysis of variance.
CUSUNORO	Cumulative Sum of Normalized Reordered Output.
DELSA	Distributed Evaluation of Local Sensitivity Analysis.
DGSM	Derivative-Based Global Sensitivity Measures.
D-scatterplot	derivative scatterplots.
FUSE	Framework for Understanding Structural Errors.
HDMR	high-dimensional model representation.
PCE	Polynomial Chaos Expansion.

References

- Anderson, B., E. Borgonovo, M. Galeotti, and R. Roson (2014), Uncertainty in climate change modelling: Can global sensitivity analysis be of help?, *Risk Anal.*, 34(2), 271–293.
- Aster, R. C., B. Borchers, and C. H. Thurber (2013), *Parameter Estimation and Inverse Problems*, Academic, Amsterdam.
- Baucells, M., and E. Borgonovo (2013), Invariant probabilistic sensitivity analysis, *Manage. Sci.*, 59(11), 2536–2549.
- Beccacece, F., and E. Borgonovo (2011), Functional ANOVA, ultramodularity and monotonicity: Applications in multiattribute utility theory, *Eur. J. Oper. Res.*, 210(2), 326–335.
- Beven, K. J. (2011), *Rainfall-Runoff Modelling: The Primer*, John Wiley, Hoboken, N. J.

Acknowledgments

We thank the Editor, Alberto Montanari, and three anonymous reviewers for several perceptive suggestions that have greatly helped us in improving our manuscript. Mary C. Hill's involvement was supported in part by the University of Kansas College of Liberal Arts and Sciences. E. Plischke acknowledges funding from German Federal Ministry of Research under grant 02S9082A. The model outputs, subroutines, and data sets can be obtained from the publicly available repository: <https://doi.org/10.5281/zenodo.885332>.

- Borgonovo, E. (2007), A new uncertainty importance measure, *Reliab. Eng. Syst. Safety*, 92(6), 771–784.
- Borgonovo, E., and E. Plischke (2016), Sensitivity analysis: A review of recent advances, *Eur. J. Oper. Res.*, 248(3), 869–887.
- Borgonovo, E., S. Tarantola, E. Plischke, and M. Morris (2014), Transformation and Invariance in the sensitivity analysis of computer experiments, *J. R. Stat. Soc., Ser. B*, 76(5), 925–947.
- Butler, M., P. Reed, K. Fisher-Vanden, K. Keller, and T. Wagener (2014a), Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis, *Environ. Modell. Software*, 59, 10–29.
- Butler, M., P. Reed, K. Fisher-Vanden, K. Keller, and T. Wagener (2014b), Inaction and climate stabilization uncertainties lead to severe economic risks, *Clim. Change*, 127(3–4), 463–474.
- Campolongo, F., J. Cariboni, and A. Saltelli (2007), An effective screening design for sensitivity analysis of large models, 22, 1509–1518.
- Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, 44, W00B02, doi:10.1029/2007WR006735.
- Clark, M. P., D. Kavetski, and F. Fenicia (2011a), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, 47, W09301, doi:10.1029/2010WR009827.
- Clark, M. P., H. K. McMillan, D. B. G. Collins, D. Kavetski, and R. A. Woods (2011b), Hydrological field data from a modeller's perspective: Part 2: Process-based evaluation of model hypotheses, *Hydrol. Processes*, 25(4), 523–543.
- Clark, M. P., et al. (2015a), A unified approach for process-based hydrologic modeling: 1. Modeling concept, *Water Resour. Res.*, 51, 2498–2514, doi:10.1002/2015WR017198.
- Clark, M. P., et al. (2015b), A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies, *Water Resour. Res.*, 51, 2515–2542, doi:10.1002/2015WR017200.
- Confalonieri, R., G. Bellocchi, S. Tarantola, M. Acutis, M. Donatelli, and G. Genovese (2010), Sensitivity analysis of the rice model WARM in Europe: Exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters, *Environ. Modell. Software*, 25(4), 479–488.
- Crnkovic, C., and J. Drachman (1996), Quality control, *Risk*, 9(9), 139–143.
- Cuntz, M., et al. (2015), Computationally inexpensive identification of noninformative model parameters by sequential screening, *Water Resour. Res.*, 51, 6417–6441, doi:10.1002/2015WR016907.
- Draper, N. R., and H. Smith (1998), *Applied Regression Analysis*, John Wiley, New York.
- Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28(4), 1015–1031.
- Efron, B., and C. Stein (1981), The jackknife estimate of variance, *Ann. Stat.*, 9(3), 586–596.
- Foglia, L., S. Mehl, M. Hill, and P. Burlando (2013), Evaluating model structure adequacy: The case of the Maggia Valley groundwater system, southern Switzerland, *Water Resour. Res.*, 49, 260–282, doi:10.1029/2011WR011779.
- Gao, L., B. A. Bryan, M. Nolan, J. D. Connor, X. Song, and G. Zhao (2016), Robust global sensitivity analysis under deep uncertainty via scenario analysis, *Environ. Modell. Software*, 76, 154–166.
- Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, 48, W08301, doi:10.1029/2011WR011044.
- Hill, M. C., and C. R. Tiedeman (2007), *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Prediction and Uncertainty*, 455 pp., John Wiley, Hoboken, N. J.
- Hill, M. C., D. Kavetski, M. Clark, M. Ye, M. Arabi, D. Lu, L. Foglia, and S. Mehl (2016), Practical use of computationally frugal model analysis methods, *Ground Water*, 54(2), 159–170.
- Homma, T., and A. Saltelli (1996), Importance measures in global sensitivity analysis of nonlinear models, *Reliab. Eng. Syst. Safety*, 52(1), 1–17.
- Hornberger, G., and R. Spear (1981), An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, 12(1), 7–18.
- Houle, E. S., B. Livneh, and J. R. Kasprzyk (2017), Exploring snow model parameter sensitivity using Sobol' variance decomposition, *Environ. Modell. Software*, 89, 144–158.
- Kuiper, N. H. (1960), Tests concerning random points on a circle, *Proc. K. Ned. Akad. Wet., Ser. A*, 63, 38–47.
- La Vigna, F., M. C. Hill, R. Rossetto, and R. Mazza (2016), Parameterization, sensitivity analysis, and inversion: An investigation using groundwater modeling of the surface-mined Tivoli-Guidonia basin (Metropolitan City of Rome, Italy), *Hydrogeol. J.*, 24(6), 1423–1441.
- Li, G., and H. Rabitz (2012), General formulation of HDMR component functions with independent and correlated variables, *J. Math. Chem.*, 50(1), 99–130.
- Lu, D., M. Ye, and M. C. Hill (2012), Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification, *Water Resour. Res.*, 48, W09521, doi:10.1029/2011WR011289.
- Marangoni, G., et al. (2017), Sensitivity of projected long-term CO₂ emissions across the shared socioeconomic pathways, *Nat. Clim. Change*, 7(2), 113–117.
- Marelli, S., and B. Sudret (2015), Uqlab user manual—Polynomial Chaos Expansions, *Tech. Rep. UQLab-V0.9–104*, Chair of Risk, Safety and Uncertainty Quant., ETH Zurich. [Available at <http://www.uqlab.com/>]
- Markstrom, S., L. Hay, and M. Clark (2016), Towards simplification of hydrologic modeling: Identification of dominant processes, *Hydrol. Earth Syst. Sci.*, 20, 4655–4671.
- Mendoza, P. A., M. P. Clark, M. Barlage, B. Rajagopalan, L. Samaniego, G. Abramowitz, and H. Gupta (2015), Are we unnecessarily constraining the agility of complex process-based models?, *Water Resour. Res.*, 51, 716–728, doi:10.1002/2014WR015820.
- Mizukami, N., M. Clark, A. Newman, A. Wood, E. Gutmann, B. Nijssen, O. Rakovec, and L. Samaniego (2017), Towards seamless large domain parameter estimation for hydrologic models, *Water Resour. Res.*, 53, doi:10.1002/2017WR020401.
- Montanari, A. (2007), What do we mean by uncertainty? The need for a consistent wording about uncertainty assessment in hydrology, *Hydrol. Processes*, 21(6), 841–845.
- Morris, M. D. (1991), Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33(2), 161–174.
- Nearing, G. S., Y. Tian, H. V. Gupta, M. P. Clark, K. W. Harrison, and S. V. Weijis (2016), A philosophical basis for hydrological uncertainty, *Hydrol. Sci. J.*, 61(9), 1666–1678.
- Norton, J. (2015), An introduction to sensitivity assessment of simulation models, *Environ. Modell. Software*, 69, 166–174.
- Paleari, L., and R. Confalonieri (2016), Sensitivity analysis of a sensitivity analysis: We are likely overlooking the impact of distributional assumptions, *Ecol. Modell.*, 340, 57–63.
- Pappenberger, F., and K. J. Beven (2006), Ignorance is bliss: Or seven reasons not to use uncertainty analysis, *Water Resour. Res.*, 42, W05302, doi:10.1029/2005WR004820.

- Pianosi, F., and T. Wagener (2015), A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, *Environ. Modell. Software*, *67*, 1–11.
- Pianosi, F., K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson, and T. Wagener (2016), Sensitivity analysis of environmental models: A systematic review with practical workflow, *Environ. Modell. Software*, *79*, 214–232.
- Plischke, E. (2010), An effective algorithm for computing global sensitivity indices (EASI), *Reliab. Eng. Syst. Safety*, *95*(4), 354–360.
- Plischke, E., E. Borgonovo, and C. Smith (2013), Global sensitivity measures from given data, *Eur. J. Oper. Res.*, *226*(3), 536–550.
- Rakovec, O., A. H. Weerts, P. Hazenberg, P. J. J. F. Torfs, and R. Uijlenhoet (2012), State updating of a distributed hydrological model with Ensemble Kalman Filtering: Effects of updating frequency and observation network density on forecast accuracy, *Hydrol. Earth Syst. Sci.*, *16*, 3435–3449.
- Rakovec, O., M. C. Hill, M. P. Clark, A. H. Weerts, A. J. Teuling, and R. Uijlenhoet (2014), Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, *Water Resour. Res.*, *50*, 409–426, doi:10.1002/2013WR014063.
- Ratto, M., P. C. Young, R. Romanowicz, F. Pappenberger, A. Saltelli, and A. Pagano (2007), Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology, *Hydrol. Earth Syst. Sci.*, *11*(4), 1249–1266.
- Razavi, S., and H. V. Gupta (2015), What do we mean by sensitivity analysis? The need for comprehensive characterization of “global” sensitivity in earth and environmental systems models, *Water Resour. Res.*, *51*, 3070–3092, doi:10.1002/2014WR016527.
- Razavi, S., and H. V. Gupta (2016), A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory, *Water Resour. Res.*, *52*, 423–439, doi:10.1002/2015WR017558.
- Rosero, E., Z.-L. Yang, T. Wagener, L. E. Gulden, S. Yatheendradas, and G.-Y. Niu (2010), Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season, *J. Geophys. Res.*, *115*, D03106, doi:10.1029/2009JD012035.
- Saltelli, A., and S. Tarantola (2002), On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal, *J. Am. Stat. Assoc.*, *97*, 702–709.
- Saltelli, A., S. Tarantola, and F. Campolongo (2000), Sensitivity analysis as an ingredient of modelling, *Stat. Sci.*, *19*(4), 377–395.
- Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008), *Global Sensitivity Analysis: The Primer*, John Wiley, Hoboken, N. J.
- Samaniego, L., et al. (2017), Toward seamless hydrologic predictions across scales, *Hydrol. Earth Syst. Sci.*, *21*, 4323–4346.
- Samuelson, P. (1941), The stability of equilibrium: Comparative statics and dynamics, *Econometrica*, *9*(2), 97–120.
- Seber, G. A. F., and C. J. Wild (1989), *Nonlinear Regression*, John Wiley, New York.
- Sobol', I. M. (1993), Sensitivity analysis for non-linear mathematical models, *Math. Modell. Comput. Exp.*, *1*, 407–414.
- Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, *55*(1–3), 271–280.
- Sobol', I., and S. Kucherenko (2009), Derivative based global sensitivity measures and their links with global sensitivity indices, *Math. Comput. Simul.*, *79*, 3009–3017.
- Sudret, B. (2008), Global sensitivity analysis using Polynomial Chaos Expansion, *Reliab. Eng. Syst. Safety*, *93*, 964–979.
- Ziehn, T., and A. S. Tomlin (2009), GUI-HDMR—A software tool for global sensitivity analysis of complex models, *Environ. Modell. Software*, *24*, 775–785.